Theory in a Virtual Observatory

Peter Teuben

Astronomy Department, University of Maryland, College Park, MD

Dave DeYoung

NOAO, Tucson, AZ

Piet Hut

Institute for Advanced Study, Princeton, NJ

Stuart Levy

National Center for Supercomputing Applications, University of Illinois Urbana-Champaign, Urbana, IL

Jun Makino

Department of Astronomy, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, JAPAN

Steve McMillan

Department of Physics and Atmospheric Science, Drexel University, Philadelphia, PA

Simon Portegies Zwart[1]

Massachusetts Institute of Technology, Cambridge, MA

Shawn Slavin

Department of Astronomy, Indiana University, Bloomington, IN

Abstract.

During the last couple of years, observers have started to make plans for a Virtual Observatory, as a federation of existing data bases, connected through levels of software that enable rapid searches, correlations, and various forms of data mining. We propose to extend the notion of a Virtual Observatory by adding archives of simulations, together with interactive query and visualization capabilities, as well as ways to simulate observations of simulations in order to compare them with observations. For this purpose, we have already organized two small workshops, earlier in 2001, in Tucson and Aspen. We have also provided concrete exam-

[1]Hubble Fellow

ples of theory data, designed to be federated with a Virtual Observatory. These data stem from a project to construct an archive for our large-scale simulations using the GRAPE-6 (a 32-Teraflops special purpose computer for stellar dynamics). We are constructing interfaces by which remote observers can observe these simulations. In addition, these data will enable detailed comparisons between different simulations.

1. Introduction

There are numerous types of theoretical data which, if integrated in a VO, will without doubt enhance its scientific capabilities. Although it has been stressed the VO itself is not intended to be a remote observatory, some branches in the theory part of a VO could very well emulate such behavior. One can imagine that after an initial selection from a set of models or a match to an observation, fine-tuning be done by re-running the models, given enough computer time and access to software (for an existing example see e.g. Pound et al. 2000).

It is perhaps instructive to view the theory part of a VO from two different points of view: that of the theorist and that of the observer.

2. The Theorist

What will a theorist find in a VO? He will find a large number of models that can be "observed". Observing such models can be done in several ways. First, one can make simulated observations of simulation data, and then compare observations with these models. Given that many models add the independent time parameter, simulations also add the complexity of exploring 4-dimensional histories and finding a best match in the time domain. The 3D spatial information will mostly likely be on a grid, or a discrete set of points.

A new and largely unused capability of theory data in a VO will be to compare models with models, much like observations are compared. This should also result in improved models, as differences and similarities between models can quickly be highlighted.

Theorists will also find a variety of standard initial conditions or benchmark data in a VO, which will make it easier to test new algorithms and compare them to previously generated data. In addition, one could also argue that besides saving the data, saving the code that generated the data will be valuable. Finally, adding theoretical data to a VO will undoubtedly also spur new data mining and CS techniques.

3. The Observer

What will an observer find about theory data in a VO? First, models can be selected and compared to observations, processing those models as though they were observed with a particular instrument.

Second, theory data can also be used to calibrate observations. Examples are: comparing Hipparchos proper motion studies with a similar analysis applied to simulations, and using stellar evolution tracks to determine cluster ages from an HRD. The added complexity of theoretical data will need new searching and matching techniques, and thus bring different type of data mining and computer science to the playing field.

4. Data Collection Toy Model

In order to develop a better understanding of theory data, we have started collecting[1] various types of theory data, mostly simulations in which time is the independent variable. Some datasets are simple benchmarks, taking initial conditions for well-known problems in Astrophysics, going back to the first published benchmark of the IAU 25-body problem (Lecar 1968).

During the IAU 208 conference in Tokyo (Teuben 2002) a survey was undertaken amongst practitioners of a well defined subset of theory data: particle simulations. These ranged from planetary to cosmological simulations, and included grid-based as well as particle-based calculations. One noteworthy find was that a surprisingly large fraction of the theorists would rather not like to see their data published in a VO, since computers get faster each year, algorithms get better and data ages quickly. Unlike observations, theoretical data often suffer from assumptions and thus comparisons can have less meaning than naively thought.

On a technical note, simulation data actually do not differ much from observational data. Most theoretical data sets fall two types: grid based ("image", each datum being the same type) or particle based (a "table" with columns and rows). An image can also be seen as a special case of a table. In recent years, added complexities are nested grids, such as in AMR, and the `tdyn` tables in Starlab's kira code (Portegies Zwart et al. 2000), where only relevant particles are updated. The `miriad` uv-data format is an example where such complexities have also been introduced to observational data. Defining the header and meta data for theoretical data will be at least as challenging as that for observational data.

5. GRAPE-6 data archive

The recently completed GRAPE-6 (Hut and Makino 1999, Makino 2002) can now produce massive datasets with a size of Terabytes for a single run. In order to handle these data, and to share them with 'guest observers', we have started to set up a data archive (see also the `manybody.org` web site). In the near future we plan to start federating our archive with other theory archives and with the budding Virtual Observatories.

Acknowledgments. We thank the Alfred P. Sloan Foundation for a grant to Hut for observing astrophysical computer simulations in the Hayden Planetarium at the Museum. We also thank the American Museum of Natural History for their hospitality.

[1] `http://www.manybody.org/`

References

Hut, P., and Makino, J. 1999, Science 283, 64

Lecar, M. 1968, Bull.Astr. 3, 91 in: Colloque sur le problem gravitationnel des N corps

Makino, J. 2002. in Astrophysical Supercomputing using Particle Simulations, ed. J. Makino, and P. Hut (San Francisco: ASP), in press

Portegies Zwart, S.F., McMillan, S.L.W., Hut, P. & Makino, J., 2001, MNRAS, 321, 199

Pound, M.W., Wolfire, M.G., Mundy, L.G., Teuben, P.J., in ASP Conf. Ser., Vol. 216, Astronomical Data Analysis Software and Systems IX, ed. N. Manset, C. Veillet, & D. Crabtree (San Francisco: ASP), 628

Teuben, P.J. 1994, in ASP Conf. Ser., Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes (San Francisco: ASP), 398

Teuben, P.J. et al. 2001, in ASP Conf. Ser., Vol. 238, Astronomical Data Analysis Software and Systems X, ed. F.R. Harnden, Jr., F.A. Primini, & H.E. Payne (San Francisco: ASP), 499

Teuben, P.J. 2002, in Astrophysical Supercomputing using Particle Simulations, ed. J. Makino, and P. Hut (San Francisco: ASP), in press